Comments on "A New Derivation of the Law of the Junctions"

Brian Hong, *Graduate Student Member, IEEE*

***Abstract*—*Contribution:* This brief comment highlights some crucial assumptions behind the "law of the junction" that are overlooked by the above paper and argues that the proposed derivation is not actually a "new" derivation at all.**

***Background:* The "law of the junction" is one of the most significant and useful results within the field of solid-state devices. The above paper is likely to confuse readers, particularly those who are undergraduate electrical engineering students studying semiconductor device physics for the first time. This is especially so because of the abstract nature of the underlying quantum mechanics framework and solid-state physics models (subjects which the typical student at that level lacks a substantial background in) as well as the plethora of tedious equations in the curriculum.**

***Research Questions:* What core physical concepts are essential to a fundamental yet intuitive understanding of the law of the junction?**

***Methodology:* Several key features of how semiconductor junctions behave under bias are explained. References to well-known textbooks are provided where appropriate.**

***Findings:* The above paper's primary mistake is its assertion that its derivation does not rely on the assumption of thermal equilibrium. However, the law of the junction is equivalent to a calculation of depletion-edge minority carrier concentrations using Maxwell–Boltzmann statistics—a distribution which only holds under thermal equilibrium conditions. More rigorously, in a nondegenerate semiconductor, Fermi–Dirac statistics (which governs electrons) reduces to Boltzmann statistics only when the electrochemical potential is spatially uniform, a condition equivalent to having no net flow of thermal energy—the very definition of thermal equilibrium.**

***Index Terms*—Electrical engineering, engineering mathematics, physical modeling, semiconductor devices.**

I. Introduction

ALTHOUGH the calculations in [1] leading to the law of the junction are correct, the accompanying intuition that is provided is dangerously flawed: "The derivation in this paper is based on counting all the electrons that have enough energy to cross from the n-side to the p-side under an externally applied voltage" [1, p. 497]. and "The actual number of electrons that cross over to the p-side depends on: 1) the density of vacant energy states in the conduction band on the p-side and 2) the number of electrons between the dashed lines on the n-side, as shown in the middle picture of [1, Fig. 2]" [1, p. 498]. No such counting argument is being used, and the aforementioned "dashed lines" make no subsequent appearance. In fact, one can see that the computation actually follows the standard, textbook procedure of directly counting the number of electrons *at the p-side depletion edge* by multiplying the density of states $N(E)$ with the Fermi–Dirac distribution $f(E)$ and integrating the product over the entire conduction band [2]–[7].[1]

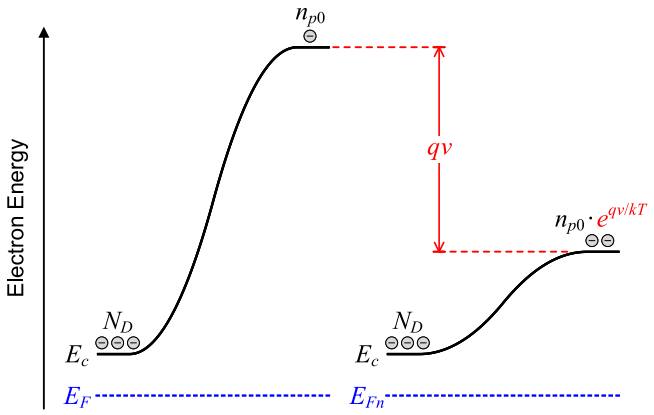

Fig. 1. Using Boltzmann statistics to deduce the law of the junction. Since the only difference between the equilibrium (left) and forward-biased (right) situations is the lowered potential energy barrier, the electron concentration at the p-side depletion edge increases by the factor $e^{qv/kT}$. Note that E_{Fn} denotes the electron quasi-Fermi level.

This comment will explain the importance of *physically* thinking about the law of the junction through the lens of thermal equilibrium—despite the fact that it is primarily deployed to *computationally* describe the behavior of junctions in nonequilibrium conditions. In particular, the next section explores in detail (based on the relationship between Maxwell–Boltzmann and Fermi–Dirac statistics) why assuming that thermal equilibrium approximately holds inside the depletion region is an indispensable step when deriving the law of the junction. Several other conceptual errors made in [1] will also be pointed out.

II. (Quasi-)Equilibrium Assumption

The above paper states that its given derivation of the law of the junction "does not make equilibrium condition

Manuscript received October 3, 2019; revised May 21, 2020, July 11, 2020, and August 31, 2020; accepted September 13, 2020.

The author was with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA. He is now with Yale Law School, Yale University, New Haven, CT 06511 USA (e-mail: bhong@caltech.edu).

Digital Object Identifier 10.1109/TE.2020.3025362

[1]While [2]–[5], being commonly used textbooks, feature standard expositions of this material, the interested reader is directed to [6] and [7] for perhaps the earliest instance of where these ideas were comprehensively developed and presented to the electrical engineering community.

assumptions" [1, Sec. I]. This is fundamentally wrong, as the given derivation assumes a constant Fermi level throughout (see [1, Fig. 2]), which by its very nature is indicative of thermal equilibrium. (Furthermore, contrary to what the band diagram in [1, Fig. 2] shows, applying a bias v to a junction actually splits the Fermi level into electron and hole *quasi*-Fermi levels, each of which must exhibit a variation of qv across the device.[2]) As discussed in numerous textbooks, the caveat that justifies relying on an equilibrium assumption in this context is that the quasi-Fermi level for a particular carrier only changes appreciably in the neutral region where the carrier is a minority carrier. In particular, therefore, the much-needed condition of a constant Fermi level throughout the depletion region, known as "quasi-equilibrium," is well-grounded [2], [6], [7, Ch. 12].[3]

It should not be surprising that the derivation in [1] relies upon the assumption of thermal equilibrium—any derivation of the law of the junction must. After all, the law of the junction is nothing more than a restatement of Maxwell–Boltzmann statistics, which is predicated on thermal equilibrium, as applied to the carriers inside the depletion region. To see this, recall that the Boltzmann distribution tells us that the concentration of particles at a particular position is proportional to the Boltzmann factor $e^{-U/kT}$, where U is the potential energy of a particle at that position [8].[4] Since an applied bias of v lowers the electrostatic potential barrier across the depletion region by v and the majority carrier concentrations are determined by the doping levels,[5] it stands to reason that the minority carrier concentrations at the edges of the depletion region are thereby enhanced by a factor of $e^{qv/kT}$. This is the law of the junction [2]–[7].

Fig. 1 portrays this line of reasoning for electrons at the two edges of the depletion region when a forward bias of v is applied. The electron concentration on the n-side is fixed by the donor doping level N_D. (Note that $n_{p0} = {n_i}^2/N_A$, although this fact is not relevant to the point being made here.) The significance of having a flat (quasi-)Fermi level throughout the depletion region—with or without bias—will become even more apparent shortly.

To convince oneself of the validity of Maxwell–Boltzmann statistics in this context, start with the Fermi–Dirac distribution $f(E)$ [given in (1) below]. Next, observe that under conditions of thermal equilibrium (i.e., constant E_F) and nondegenerate doping (i.e., $E - E_F$ larger than several kT for $E \geq E_c$), $f(E)$ becomes proportional to the Boltzmann factor $e^{-E/kT}$. Consequently, changing the potential energy [i.e., shifting the entire band structure or density of states $N(E)$ along the energy axis] by ΔE enhances the occupation probability of each state—and hence the electron concentration—by the factor $e^{-\Delta E/kT}$. This completes the argument.[6] An analogous discussion holds for holes.

Fig. 2 provides an illustrative example of the band diagram of an abrupt pn-junction, both in equilibrium and with an applied bias.[7] As one can see, the quasi-Fermi level for each carrier remains relatively flat inside the depletion region and only experiences a significant variation in the neutral region where the carrier is a minority carrier. Of course, neither quasi-Fermi level is ever *truly* flat under bias (at any position inside the device), otherwise there would be no net current for that carrier (at that position)—in violation of Kirchhoff's Current Law (KCL) [9]. But, the quasi-equilibrium assumption is justified by the fact that only a comparatively tiny gradient in the Fermi level is needed to sustain the required current when the concentration of carriers is high.[8]

Using the law of the junction, one can verify that the electron and hole concentrations under bias at the p- and n-side depletion edges are 5.26×10^{12} cm^{-3} and 2.10×10^{12} cm^{-3}, respectively. Note that calculation of the current densities would require the diffusion coefficients (or mobilities) of both carriers. The distinction between short and long diodes is highlighted for educational purposes—specifically, to showcase the difference between their (minority-carrier) quasi-Fermi level profiles. A "short" junction is one whose physical length is much smaller than the diffusion lengths of the carriers. This causes practically all of the excess minority carriers to recombine at the contacts and their concentrations to therefore vary linearly with position in the neutral regions. Both "short" and "long" junctions find applications in numerous settings: The former are characteristic of most of the transistors in modern silicon/SiGe integrated circuits, while the latter are prevalent in

[2]Recall that a voltage source or battery is ultimately an *electrochemical* cell. In a semiconductor, electrochemical potential and Fermi level are synonymous.

[3]This subtle but important point is unfortunately glossed over in some textbooks (e.g., [5]).

[4]Combining the Boltzmann distribution with the fact that drift and diffusion cancel in thermal equilibrium, one can derive the Einstein relation, which relates the mobility of the particles to their diffusion coefficient.

[5]Note that these two claims are valid only for *low-level injection*, where the applied forward bias is low enough that the concentrations of injected excess minority carriers are small in comparison to the background doping densities—that is, the majority carriers are still the *overwhelming* majority in their respective neutral regions. Otherwise, if the injection level into a particular side is *high*, the majority carriers on that side will increase beyond their thermal equilibrium amount (which is dictated by the doping level), causing part of the applied bias to appear across that side's neutral region. In this case, one would resort to a more general form of the law of the junction, $np = {n_i}^2 e^{qv/kT}$ (which, along with charge neutrality, yields the carrier concentrations) [2]–[5], although a detailed analysis of this scenario is beyond the scope of this article.

[6]This can be shown rigorously. Keeping in mind that $f(E) \propto e^{-E/kT}$,

$$\begin{aligned} n_{\text{final}} &= \int_{E_c+\Delta E}^{\infty} N(E-\Delta E)\, f(E)\, dE \\ &= \int_{E_c}^{\infty} N(E')\, f(E'+\Delta E)\, dE' \\ &= e^{-\Delta E/kT} \int_{E_c}^{\infty} N(E')\, f(E')\, dE' \\ &= n_{\text{initial}} \cdot e^{-\Delta E/kT}. \end{aligned}$$

[7]Note that the values for the intrinsic carrier concentration n_i, the band gap E_g, and the relative permittivity $\varepsilon_r = 11.8$ are borrowed from silicon, although the discussion here does not depend on the material. The portrayed band diagrams remain conceptually representative so long as the junction has the same material on both sides (as opposed to a heterojunction).

[8]Another way of looking at the quasi-equilibrium assumption is by noting that drift and diffusion in the depletion region *roughly* cancel, resulting in a net current that is significantly smaller than the individual drift and diffusion components of that current (often by several orders of magnitude). Thus, the depletion region's electrochemical gradient, which controls the net current, must be considerably weaker than the electric field (i.e., the band bending), which causes drift.

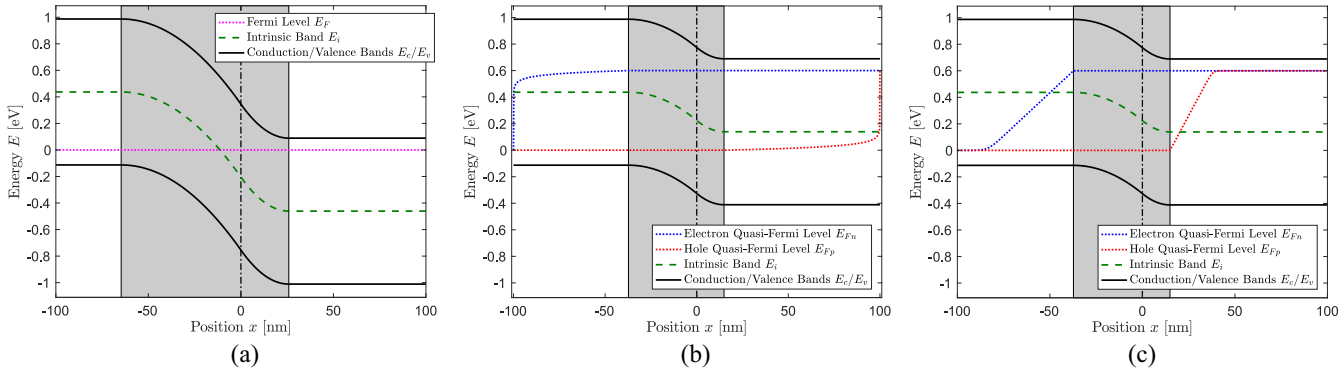

Fig. 2. Band diagram of an abrupt pn-junction with doping levels $N_A = 2 \times 10^{17}$ cm^{-3} and $N_D = 5 \times 10^{17}$ cm^{-3}, an intrinsic carrier concentration of $n_i = 10^{10}$ cm^{-3}, and a band gap of $E_g = 1.1$ eV. The thermal voltage is taken to be $kT/q = 26$ mV. The built-in potential is therefore $\psi_0 = 0.898$ V. The p- and n-side contacts are at $x = \mp 100$ nm, respectively, and the metallurgical junction between the two sides is at $x = 0$. The Fermi level at the p-side contact is referenced to 0. The depletion region has been shaded. (a) Equilibrium (no bias). (b) Forward bias of 0.6 V. The minority carrier diffusion lengths are $L_p = 10$ μm and $L_n = 20$ μm, resulting in a "short" diode. (c) Forward bias of 0.6 V, but with $L_p = 1$ nm and $L_n = 2$ nm, resulting in a "long" diode.

photonic devices made from direct band gap semiconductors (such as GaAs or InP laser diodes) due to the shorter lifetimes of the carriers [5].[9]

III. Fermi–Dirac Distribution

In the middle of the 2nd column of [1, p. 497], the above paper erroneously refers to the Fermi–Dirac distribution $f(E)$, given in (1), as a probability density function (pdf):

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}}. \tag{1}$$

Recall that for the general case of a random variable $X \in \mathbb{R}^n$, its pdf is a mapping $f_X \colon \mathbb{R}^n \to \mathbb{R}^+$ such that for any subset $S \subseteq \mathbb{R}^n$, the probability that $X \in S$ is given by f_X integrated over S [10]. That is, $\mathbb{P}(X \in S) = \int_S f_X(x)\, dx$. As such, the units of f_X are the *reciprocal* of the units of X. In particular, f_X must also integrate to unity over all of $\mathbb{R}^n$.

In this case, for *each* value of E, the Fermi–Dirac distribution $f(E)$ is itself a probability. Specifically, $f(E)$ is the probability that a state at energy E is occupied by a fermion, and therefore gives the average fraction of (single-fermion) states at energy E that are occupied. Notice further that $f(E)$, being unit-less, does not have the right units to be a pdf over energy. Thus, one can clearly see that the Fermi–Dirac distribution $f(E)$ is *not* a pdf.

As a result, the equations [1, eqs. (1) and (2)] are not to be interpreted as "integrating $N(E)$ against $f(E)\, dE$" as one would do if f were a pdf, but rather as "integrating the (occupied) density of states $N(E)\, f(E)$ over E." The density of states $N(E)$ is defined as the partial derivative of the total number of states (per unit volume) up to the energy level E with respect to E. As such, the product $N(E)\, f(E)$ gives the rate at which the total number of *occupied* states (per unit volume) up to energy E increases with E.

[9]Note that a "long" diode can be used to approximate the hypothetical scenario where minority carriers are injected into a semi-infinitely long semiconductor slab. In this situation, the minority carrier's quasi-Fermi level changes linearly with distance at a slope given by kT/L (where L is the minority carrier's diffusion length) until it "meets" the quasi-Fermi level of the majority carrier, at which point effectively all of the excess minority carriers will have recombined.

IV. Applied Forward Bias

The first paragraph on [1, p. 498] states that "in practical device operation," the externally applied voltage across a junction, v, is "much smaller than" the built-in potential, ψ_0.[10] This is patently incorrect. What the author is most likely referring to is the *small-signal voltage* that is superimposed upon the operating-point bias voltage.[11] This bias voltage, by contrast, must "turn on" the junction and is typically a substantial fraction of ψ_0. To provide the reader with a correct numerical feel for these quantities in modern (discrete) microelectronic applications, ψ_0 is roughly 1 V for doping levels on the order of 10^{18} cm^{-3}, while the base-emitter turn-on voltage of a standard 2N3904 transistor is typically around 650–700 mV (for milliampere-level currents). Therefore, the applied voltage v, which is the *sum* of the bias and small-signal voltages, is not "in the range of 5–20 mV" as [1] states.

V. Pedagogical Conclusion

Although seeking out and bringing to light new perspectives into existing analyses and computations is undoubtedly a pedagogically valuable exercise, one must be careful to truly understand the elementary principles in the underlying models. In this particular context, for example, it is important to realize that no derivation of the law of the junction can forgo the quasi-equilibrium assumption of the (quasi-)Fermi level(s) (for both carriers) being constant inside the junction's depletion region. This is because the law of the junction is essentially Maxwell–Boltzmann statistics applied to the setting of a pn-junction, and this distribution is applicable only when the particles are in a state of thermal equilibrium.

[10]Recall that the built-in potential (also known as the contact potential) is the electrostatic potential difference that develops across a junction in thermal equilibrium due to the electric field generated by the ionized acceptor and donor dopant atoms inside the depletion region (also called the space charge or transition region). By facilitating a drift current that offsets the diffusion of carriers between the junction's two sides, the built-in potential is a thermodynamic manifestation of the junction reaching thermal equilibrium.

[11]Of course, the signal voltage need not be small. Small signals are typically used when the linear operation of the device is desired. But, there are many applications where this is not the case (e.g., the transistors in a switching power amplifier).

REFERENCES

[1] H. R. Pota, "A new derivation of the law of the junctions," *IEEE Trans. Educ.*, vol. 47, no. 4, pp. 497–499, Nov. 2004.

[2] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, NJ, USA: Wiley, 2007.

[3] C. C. Hu, *Modern Semiconductor Devices for Integrated Circuits*, 1st ed. Upper Saddle River, NJ, USA: Pearson Educ., Inc., 2009. [Online]. Available: https://www.chu.berkeley.edu/modern-semiconductor-devices-for-integrated-circuits-chenming-calvin-hu-2010/

[4] B. V. Zeghbroeck, *Principles of Semiconductor Devices*, 2011. [Online]. Available: http://ecee.colorado.edu/∼bart/book/book/title.htm

[5] B. G. Streetman and S. K. Banerjee, *Solid State Electronic Devices*, 7th ed. Upper Saddle River, NJ, USA: Pearson Educ., Inc., 2014.

[6] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Syst. Tech. J.*, vol. 28, no. 3, pp. 435–489, Jul. 1949.

[7] W. Shockley, *Electrons and Holes in Semiconductors With Applications to Transistor Electronics*. Princeton, NJ, USA: D. Van Nostrand Company, Inc., 1950.

[8] F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion*, 3rd ed. Cham, Switzerland: Springer Int., 2016.

[9] M. E. V. Valkenburg, *Network Analysis*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1974.

[10] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Upper Saddle River, NJ, USA: Pearson Educ., Inc., 2008.

Brian Hong (Graduate Student Member, IEEE) received the B.S. degree (*summa cum laude*) in electrical engineering from the University of California at Los Angeles, Los Angeles, CA, USA, in 2013, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2014 and 2018, respectively. He is currently pursuing the J.D. degree with Yale Law School, New Haven, CT, USA.

His research interests include the mathematical and physical modeling of electronic systems, with a particular emphasis on the theory of injection locking and pulling in electrical oscillators.

Dr. Hong was a recipient of the JPL Undergraduate Scholarship in 2008, the Rose Hills Foundation Fellowship in 2013, the Analog Devices Outstanding Student Designer Award in 2014, and the William S. Beinecke Scholarship in 2019. He currently serves as an Editor of the *Yale Law Journal*.